

Requested Patent: JP2001216098A
Title: DISK ARRAY DEVICE AND DISK CONTROL METHOD ;
Abstracted Patent: JP2001216098 ;
Publication Date: 2001-08-10 ;
Inventor(s): SHIMIZU KUNIYASU ;
Applicant(s): TOSHIBA CORP ;
Application Number: JP20000022659 20000131 ;
Priority Number(s): ;
IPC Classification: G06F3/06; G06F12/16 ;
Equivalents: ;

ABSTRACT:

PROBLEM TO BE SOLVED: To provide a disk array device which can prevent data loss even to the second disk fault occurring during a degenerate operation. **SOLUTION:** This disk array device 15 is a logical drive including plural disk devices 152 arranged in parallel to each other and can improve its fault resistance by using the redundant data for reproducing the data on a certain device 152 by means of the data on another devices 152. When a degenerate operation is carried out due to the failure of one of devices 152, a host bus adapter(HBA) 151 controlling the device 15 secures one or more idle device 152 among other devices 152 and then reconfigures the new redundant data to reproduce the remaining effective data by using those idle devices 152.

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2001-216098
(P2001-216098A)

(43)公開日 平成13年8月10日(2001.8.10)

| (51)Int.Cl. ⁷ | 識別記号 | F I | テ-リ-ト*(参考) |
|--------------------------|-------|--------------|-------------------|
| G 0 6 F 3/06 | 5 4 0 | G 0 6 F 3/06 | 5 4 0 5 B 0 1 8 |
| | 3 0 1 | | 3 0 1 M 5 B 0 6 5 |
| 12/16 | 3 2 0 | 12/16 | 3 2 0 L |

審査請求 未請求 請求項の数13 O L (全 10 頁)

(21)出願番号 特願2000-22659(P2000-22659)

(22)出願日 平成12年1月31日(2000.1.31)

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 清水 邦保

東京都青梅市末広町2丁目9番地 株式会
社東芝青梅工場内

(74)代理人 100058479

弁理士 鈴江 武彦 (外6名)

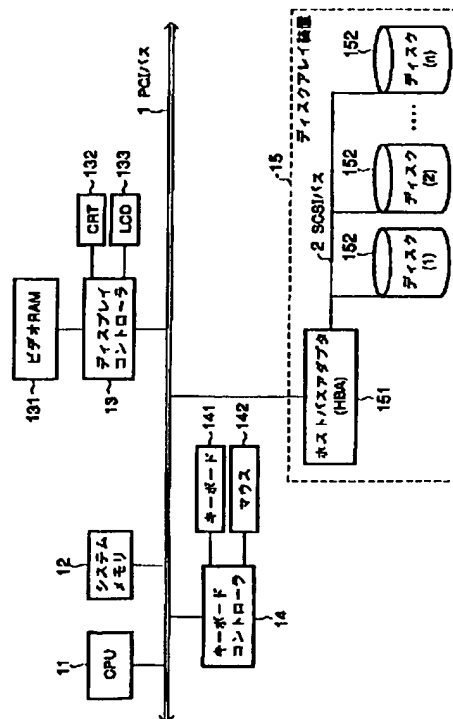
Fターム(参考) 5B018 GA06 HA24 HA35 KA15 MA14
RA115B065 BA01 CA12 CA30 CC04 CC08
CS04 EA02 EA12 EA18

(54)【発明の名称】 ディスクアレイ装置およびディスク制御方法

(57)【要約】

【課題】縮退運転中におけるさらなるディスク障害に対しても、データ損失を防ぐことを可能とするディスクアレイ装置を提供する。

【解決手段】ディスクアレイ装置15は、複数のディスク装置152を並列配置した論理的なドライブであり、いずれかのディスク装置152上のデータを他のディスク装置152上のデータを用いて再生するための冗長データをもつことにより耐障害性を向上させている。そして、このディスクアレイ装置15を制御するホストバスアダプタ(HBA)151は、いずれかのディスク装置152が故障したことにより縮退運転に移行したときに、残りのディスク装置152の中から空のディスク装置152を1つ以上確保し、この確保した空のディスク装置152を用いて、残存する有効データを再生するための新たな冗長データをさらに構築する。



【特許請求の範囲】

【請求項1】 物理的に異なる複数のディスク装置を並列配置して論理的に単一のディスク装置を構成し、あるディスク装置上のデータを他のディスク装置上のデータを用いて再生するための冗長データを持つことにより耐障害性を向上させたディスクアレイ装置において、いずれかのディスク装置が障害を発生させたことにより前記冗長データを用いた縮退運転に移行したときに、その他のディスク装置の中から有効データを持たない空きのディスク装置を1つ以上確保し、前記その他のディスク装置上の前記冗長データを含む有効データを再生するための新たな冗長データを前記確保した空きのディスク装置上に構築するディスク制御手段を具備することを特徴とするディスクアレイ装置。

【請求項2】 前記ディスク制御手段は、前記その他のディスク装置上の有効データを圧縮することにより前記空きのディスク装置を確保することを特徴とする請求項1記載のディスクアレイ装置。

【請求項3】 前記ディスク制御手段は、前記有効データの圧縮前および圧縮後の格納位置の対応関係をアドレス変換テーブルに記録することを特徴とする請求項2記載のディスクアレイ装置。

【請求項4】 前記ディスク制御手段は、前記アドレス変換テーブルを定期的に走査して、前記その他のディスク装置の空き領域が連続するように前記圧縮後の有効データを再配置する再配置手段を具備することを特徴とする請求項3記載のディスクアレイ装置。

【請求項5】 前記ディスク制御手段は、書き込み処理時に空き領域不足が検知されたときに、前記アドレス変換テーブルを走査して、前記その他のディスク装置の空き領域が連続するように前記圧縮後の有効データを再配置する再配置手段を具備することを特徴とする請求項3記載のディスクアレイ装置。

【請求項6】 前記再配置手段は、仮想アドレス順となるように前記圧縮後の有効データを再配置することを特徴とする請求項4または5記載のディスクアレイ装置。

【請求項7】 前記ディスク制御装置は、前記その他のディスク装置の空き領域が連続するように前記有効データを移動させることにより前記空きのディスク装置を確保することを特徴とする請求項1記載のディスクアレイ装置。

【請求項8】 前記ディスク制御手段は、前記有効データの移動前および移動後の格納位置の対応関係をアドレス変換テーブルに記録することを特徴とする請求項7記載のディスクアレイ装置。

【請求項9】 前記ディスク制御手段は、仮想アドレス順となるように前記有効データを移動させることを特徴とする請求項8記載のディスクアレイ装置。

【請求項10】 前記ディスク制御手段は、前記その他のディスク装置の空き領域が連続するように前記有効データ

ータを移動させた後、この移動後の有効データを圧縮することにより前記空きのディスク装置を確保することを特徴とする請求項1記載のディスクアレイ装置。

【請求項11】 物理的に異なる複数のディスク装置を並列配置して論理的に単一のディスク装置を構成し、あるディスク装置上のデータを他のディスク装置上のデータを用いて再生するための冗長データを持つことにより耐障害性を向上させたディスクアレイ装置のディスク制御方法において、

いずれかのディスク装置が障害を発生させたことにより前記冗長データを用いた縮退運転に移行したときに、その他のディスク装置上の有効データを圧縮することにより有効データを持たない空きのディスク装置を1つ以上確保し、前記その他のディスク装置上の前記冗長データを含む有効データを再生するための新たな冗長データを前記確保した空きのディスク装置上に構築することを特徴とするディスク制御方法。

【請求項12】 物理的に異なる複数のディスク装置を並列配置して論理的に単一のディスク装置を構成し、あるディスク装置上のデータを他のディスク装置上のデータを用いて再生するための冗長データを持つことにより耐障害性を向上させたディスクアレイ装置のディスク制御方法において、

いずれかのディスク装置が障害を発生させたことにより前記冗長データを用いた縮退運転に移行したときに、その他のディスク装置上の空き領域が連続するように有効データを移動させることにより有効データを持たない空きのディスク装置を1つ以上確保し、前記その他のディスク装置上の前記冗長データを含む有効データを再生するための新たな冗長データを前記確保した空きのディスク装置上に構築することを特徴とするディスク制御方法。

【請求項13】 物理的に異なる複数のディスク装置を並列配置して論理的に単一のディスク装置を構成し、あるディスク装置上のデータを他のディスク装置上のデータを用いて再生するための冗長データを持つことにより耐障害性を向上させたディスクアレイ装置のディスク制御方法において、

いずれかのディスク装置が障害を発生させたことにより前記冗長データを用いた縮退運転に移行したときに、その他のディスク装置上の空き領域が連続するように有効データを移動させた後、この移動後の有効データを圧縮することにより有効データを持たない空きのディスク装置を1つ以上確保し、前記その他のディスク装置上の前記冗長データを含む有効データを再生するための新たな冗長データを前記確保した空きのディスク装置上に構築することを特徴とするディスク制御方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、たとえばRAID

D技術等による高信頼性のディスクアレイ装置および同装置のディスク制御方法に係り、特に、縮退運転中におけるさらなるディスク障害に対しても、データ損失を防ぐことを可能とするディスクアレイ装置およびディスク制御方法に関する。

【0002】

【従来の技術】近年、様々な分野で業務の電子化が急速に進んでおり、データ処理に対する信頼性は高レベルのものが要求されている。また、このような状況から、ディスクの耐障害性に対する要求も、日々高まっていく一方である。そして、冗長データを持つことにより、ディスクの耐障害性を向上させる方法として、RAID技術が知られている。

【0003】特に、RAID5は、論理ドライブ(アレイ)を構成する複数のディスク装置の中のいずれか1つのディスク装置が故障した場合であっても、冗長データを用いることにより、その故障したディスク装置上のデータを再生することが可能であるため、非常に高い耐障害性を実現している。

【0004】

【発明が解決しようとする課題】ところで、これまでのRAIDシステムでは、冗長データを持ったディスクアレイであっても、あるディスク装置が故障したことにより縮退運転になると、この縮退運転中は、残りのディスク装置のいずれかが故障しても、データが損失してしまうといった問題があった。

【0005】たとえば、RAID5を構成するディスクアレイでは、ディスク装置が1つ故障した場合に、そのままの状態でも縮退運転によりデータアクセスを継続することが可能であるが、他のいずれかのディスク装置がさらに故障した場合には、即時にデータ損失に至ってしまう。

【0006】この発明はこのような事情を考慮してなされたものであり、縮退運転中におけるさらなるディスク障害に対しても、データ損失を防ぐことを可能とするディスクアレイ装置およびディスク制御方法を提供することを目的とする。

【0007】

【課題を解決するための手段】前述した目的を達成するために、この発明のディスクアレイ装置は、アレイを構成する複数のディスク装置の中のいずれかが故障したことにより縮退運転に移行したときに、残りのディスク装置の中から空のディスク装置を確保し、この確保した空のディスク装置を用いて、残存する有効データを再生するための新たな冗長データをさらに構築するようにしたものであり、そのために、物理的に異なる複数のディスク装置を並列配置して論理的に単一のディスク装置を構成し、あるディスク装置上のデータを他のディスク装置上のデータを用いて再生するための冗長データを持つことにより耐障害性を向上させたディスクアレイ装置にお

いて、いずれかのディスク装置が障害を発生させたことにより前記冗長データを用いた縮退運転に移行したときに、その他のディスク装置の中から有効データを持たない空のディスク装置を1つ以上確保し、前記その他のディスク装置上の前記冗長データを含む有効データを再生するための新たな冗長データを前記確保した空のディスク装置上に構築するディスク制御手段を具備することを特徴とする。

【0008】そして、残りのディスク装置の中から空のディスク装置を確保するために、この発明のディスクアレイ装置は、前記ディスク制御手段が、(1)前記その他のディスク装置上の有効データを圧縮し、(2)前記その他のディスク装置の空き領域が連続するように前記有効データを移動させ、または、(3)前記その他のディスク装置の空き領域が連続するように前記有効データを移動させた後、この移動後の有効データを圧縮する、等を行うようにしたものである。

【0009】この発明のディスクアレイ装置においては、縮退運転に移行した際、残存する有効データ(故障したディスク装置上のデータを再生するための冗長データを含む)を再生するための新たな冗長データをさらに構築するため、この縮退運転中にいずれかのディスク装置がさらに故障した場合であっても、新たに構築した冗長データを用いて縮退運転移行時に残存した有効データすべてを保証でき、これにより、通常稼働時のすべてのデータを保証できることになる。

【0010】また、この発明のディスクアレイ装置は、前記有効データの圧縮前および圧縮後の格納位置の対応関係、または、前記有効データの再配置前および再配置後の格納位置の対応関係をアドレス変換テーブルに記録し、定期的または所定のイベント発生時に、このアドレス変換テーブルに基づき、前記その他のディスク装置の空き領域が連続するように、さらには、仮想アドレス順となるように、前記有効データを再配置することが好ましい。

【0011】この発明のディスクアレイ装置においては、縮退運転中のディスクアレイに対して冗長データを再構成できるため、さらにディスク装置が故障した場合でもアレイ上のデータを失うことなくデータアクセスすることができる。また、領域を仮想アドレス順に配置することにより、故障していないディスクへのアクセス性能を向上することができる。

【0012】

【発明の実施の形態】以下、図面を参照してこの発明の実施形態を説明する。図1は、この発明の実施形態に係るディスクアレイ装置を適用するコンピュータシステムの構成図である。

【0013】図1に示すように、このコンピュータシステムでは、システムバスとしてPCIバス1が敷設されており、このPCIバス1に、CPU11、システムメ

モリ12、ディスプレイコントローラ13、キーボードコントローラ14およびホストバスアダプタ(HBA)151が接続されている。

【0014】CPU11は、このコンピュータシステム全体の制御を司るものであり、システムメモリ12に格納されたオペレーティングシステムやユーティリティを含むアプリケーションプログラムの記述に基づいて動作する。

【0015】システムメモリ12は、このコンピュータシステムの主記憶となるメモリデバイスであり、CPU11の動作を記述したオペレーティングシステムやユーティリティを含むアプリケーションプログラムを格納する。

【0016】ディスプレイコントローラ13は、このコンピュータシステムのユーザインタフェースのアウトプットを司るものであり、CPU11により描画されたビデオRAM131内の表示データをCRT132やLCD133等の表示装置に表示制御する。

【0017】キーボードコントローラ14は、このコンピュータシステムのユーザインタフェースのインプットを司るものであり、キーボード141やマウス142等の入力装置の操作をデータ化してCPU11に引き渡す。

【0018】ホストバスアダプタ(HBA)151は、このコンピュータシステムの外部記憶となるディスクアレイ装置15を構成するために、SCSIバス2により接続される複数のディスク装置152を制御するものである。このホストバスアダプタ(HBA)151により構成されるディスクアレイ装置15は、複数のディスク装置152が並列配置された論理的なドライブであり、いずれかのディスク装置152上のデータを他のディスク装置上のデータを用いて再生するための冗長データをもつことにより、耐障害性を向上させている。

【0019】そして、この実施形態のディスクアレイ装置15は、いずれかのディスク装置152の故障により冗長データを用いた縮退運転に移行している最中に、さらなるディスク装置152の故障が発生したときであっても、データ損失を防止してディスクアクセスを継続可能とした点を特徴としており、以下、この点について詳述する。

【0020】(第1実施形態) まず、このディスクアレイ装置15におけるディスク制御の第1実施形態について説明する。図2は、この第1実施形態に係るホストバスアダプタ(HBA)151の構成を示す図である。

【0021】図2に示すように、このホストバスアダプタ(HBA)151は、アレイコントローラ1511、SCSI制御機構1512、故障回復機能1513およびアドレス変換テーブル1514を備えている。

【0022】アレイコントローラ1511は、このディスクアレイ装置15の中核をなすものであり、通常稼働

時におけるデータの配置制御や冗長データの生成を実行するのみならず、縮退運転時におけるデータ再生および以下に示すこの発明に特有のディスク制御を実行する。

【0023】すなわち、アレイコントローラ1511は、図3に示すように、アレイを構成するディスクに故障が発生したときに(図3(a))、その他のディスク上のデータを圧縮することにより空のディスクを確保し(図3(b))、この空のディスクを用いてその他のディスク上のデータに関する冗長データをさらに構築する(図3(c))、といったディスク制御を実行する。

【0024】また、SCSI制御機構1512は、SCSIバス2で定められたプロトコルに基づき、ディスク装置152との間のデータ通信を実行する。故障回復機構1513は、故障したディスク装置152が新たなディスク装置152と交換されたときに、アレイコントローラ1511とは独立して、この交換後のディスク装置152が通常稼働時にもつべきデータを再構築する。そして、アドレス変換テーブル1514は、縮退運転時におけるこの発明に特有のディスク制御を実現するためのアドレス情報を記録する。

【0025】アレイコントローラ1511は、縮退状態では稼働中のアレイ(N個のディスク装置)のデータを圧縮して、アレイ上のすべての有効なデータをN-1個のディスク上に置く。そして、この有効なデータを持つN-1個のディスク装置上のデータのための冗長データを有効なデータを持たない1個のディスク上に再構成する。圧縮後のデータ容量がN-1個のディスクでは不足している場合には、冗長データを再構成せず、縮退モードのまま稼働する。

【0026】この圧縮は、一定のデータブロック(セクタなど)単位で行う。データブロックの圧縮後は元のデータサイズよりも小さくなるが、図4に示すように、アレイ上ではデータブロックサイズアラインしないで領域を割り当てる。このため、ファイルシステムが一つのデータブロックをアクセスする場合でも、実際にアレイ上では、複数の(連続する)データブロックに圧縮されたデータが存在する場合がある。

【0027】そして、この圧縮時、ディスク制御装置はファイルシステムから見えるデータブロックアドレスと実際のアレイ上のデータブロックとの対応を、前述のアドレス変換テーブル1514に記録する。

【0028】ホスト側からディスクの読み出しが要求された場合、アレイコントローラ1511は、アドレス変換テーブル1514を参照しながら圧縮されたデータを読み出した後、この読み出したデータを展開し、この展開後のデータをホスト側に引き渡す。

【0029】一方、ホスト側からディスクへの書き込みが要求された場合には、アレイコントローラ1511は、まず、アドレス変換テーブル1514を参照しながらこの要求された書き込みブロックに相当する古い圧縮

データがディスクアレイ上に存在するかどうかを判定し、存在した場合には、この古いデータを破棄し、このデータが記録されていたブロックを空き領域とする。

【0030】次に、要求されたデータを圧縮してディスクアレイ上の空き領域に書き込み、新たに書き込んだ位置をアドレス変換テーブル1514に記録する。図5は、この書き込みによるブロックの移動を示す図である。

【0031】また、アレイコントローラ1511は、定期的に、または、圧縮した後のデータを書き込むだけの領域がディスクアレイ上に残っていないときに、図6に示すように、アレイ上の空き領域が連続するようにデータ領域の再割り当て（ガベージコレクト）を実行する。このとき、さらに、図7に示すように、仮想アドレス順となるように考慮しながらデータ領域の再割り当て（ガベージコレクト）を実行しても良い。

【0032】なお、ホスト側から書き込みが要求された際、再割り当てを行ったにも拘わらずに、圧縮後のデータを書き込むための領域が新たにアレイ上に確保できなかった場合には、ホスト側に書き込み要求が失敗したことを通知する。

【0033】これにより、縮退運転時、データの圧縮により新たに冗長ディスクを確保できさえすれば、さらにもう1つのディスク故障が発生した場合であっても、ディスクアクセスを継続することが可能となる。

【0034】次に、図8乃至図10を参照して、このディスクアレイ装置15の動作手順を説明する。

【0035】ディスク装置152の故障発生時、アレイコントローラ1511は、故障していないディスク装置152が2つ以上存在するかどうか判定し（図8のステップA1）、存在していれば（図8のステップA1のYES）、これらの有効データを圧縮しながら、アドレス変換テーブル1514を構成する（図8のステップA2）。

【0036】次に、アレイコントローラ1511は、この圧縮により有効データの無いディスク装置152が確保できたかどうか判定し（図8のステップA3）、確保できていれば（図8のステップA3のYES）、その有効データの無いディスク装置152上にその他のディスク装置152上のデータに関する冗長データを構成する（図8のステップA4）。

【0037】また、縮退運転中における読み出しリクエスト発生時、アレイコントローラ1511は、そのデータが圧縮済みかどうかを判定し（図9のステップB1）、圧縮済みでなければ（図9のステップB1のNO）、要求されたアドレスからデータを読み出す（図9のステップB2）。

【0038】一方、圧縮済みであれば（図9のステップB1のYES）、アレイコントローラ1511は、アドレス変換テーブルを参照して圧縮されたデータの位置を

確定し（図9のステップB3）、圧縮されたデータを読み出して展開し（図9のステップB4）、この展開したデータの中から要求されたデータをホスト側に引き渡す（図9のステップB5）。

【0039】また、縮退運転中における書き込みリクエスト発生時、アレイコントローラ1511は、そのデータが圧縮済みかどうかを判定し（図10のステップC1）、圧縮済みでなければ（図10のステップC1のNO）、要求されたアドレスにデータを書き込む（図10のステップC2）。

【0040】一方、圧縮済みであれば（図10のステップC1のYES）、アレイコントローラ1511は、この書き込みデータを圧縮し（図10のステップC3）、この圧縮後のデータを書き込むだけの空き領域が存在するかどうか判定する（図10のステップC4）。

【0041】ここで、空き領域がなければ（図10のステップC4のNO）、アレイコントローラ1511は、データ領域の再割り当て（ガベージコレクト）を実行し（図10のステップC5）、空き領域が確保されたかどうか判定する（図10のステップC6）。そして、空き領域が確保されなかった場合には（図10のステップC6のNO）、この書き込みリクエストをエラー終了する。

【0042】また、空き領域があった場合（図10のステップC4のYES）および再割り当て（ガベージコレクト）により空き領域が確保された場合（図10のステップC6のYES）、アレイコントローラ1511は、アドレス変換テーブルを参照してアレイ上の対応する旧圧縮データを破棄するとともに、新たに圧縮したデータをアレイ上の空き領域に書き込んでアドレス変換テーブルを更新する（図10のステップC7）。

【0043】以上のように、このディスクアレイ装置15は、縮退運転中におけるさらなるディスク障害に対しても、データ損失を防ぐことを可能とする。

【0044】（第2実施形態）次に、このディスクアレイ装置15におけるディスク制御の第2実施形態について説明する。

【0045】図11は、この第2実施形態に係るホストバスアダプタ（HBA）151の構成を示す図である。

【0046】図11に示すように、このホストバスアダプタ（HBA）151は、前述の第1実施形態において図2に示したホストバスアダプタ（HBA）151の構成に加えて、有効ブロックテーブル1515を新たに設けたものである。この有効ブロックテーブル1515は、有効データが保存されている領域を記録するために用いる。また、この第2実施形態では、アドレス変換テーブル1514を有効データの圧縮前と圧縮後の対応関係を記録するためではなく、後述する有効データの移動前と移動後の対応関係を記録するために用いる。

【0047】そして、この第2実施形態では、アレイコ

ントローラ1511は、図12示すように、アレイを構成するディスクに故障が発生したときに(図12

(a))、その他のディスク上のデータを移動させることにより空のディスクを確保し(図12(b))、この空のディスクを用いてその他のディスク上のデータに関する冗長データをさらに構築する(図12(c))、といったディスク制御を実行する。

【0048】アレイコントローラ1511は、ディスクアレイの初期化時以降のディスク書き込みを監視し、アレイ上でファイルシステムに使用された経歴のある領域(有効なデータが保存されているとみなす)をすべて有効ブロックテーブル1515に記録する。

【0049】そして、縮退モードに入った(N個のディスク装置で稼動中)場合、有効ブロックテーブルに記録されたすべてのブロック上のデータを(N-1)個の稼動中のディスク装置上に移動し、その移動前と移動後との対応を、前述のアドレス変換テーブル1514に記録する。これにより、有効なデータをもたないディスク装置152を1つ以上確保する。

【0050】有効なデータを持たないディスク装置が確保できれば、ここに残りのN-1個のディスク装置上のデータのための冗長データを再構成する。もしN-1個のディスクでは容量が不足している場合には、冗長データを再構成せず、縮退モードのまま稼動する。

【0051】ホスト側から読み出しが要求された場合、アレイコントローラ1511は、アドレス変換テーブル1514を参照してアレイ上のデータの位置を特定し、この要求されたデータを読み出す。一方、ホスト側から書き込みが要求された場合、アレイコントローラ1511は、アドレス変換テーブル1514に登録されたブロックへの書き込みについてはアドレス変換テーブル1514で指定された領域に書き込み、また、アドレス変換テーブル1514に登録されていないブロックへの書き込みについてはN-1個のディスク装置上の領域から有効でないブロックを新たに確保し、アドレス変換テーブル1514に登録するとともに、この領域に指定されたデータを書き込む。なお、新たな領域が確保できなかった場合には、ホスト側に書き込み要求が失敗したことを通知する。

【0052】これにより、縮退運転時、データの移動により新たに冗長ディスクを確保できさえすれば、さらにもう1つのディスク故障が発生した場合であっても、ディスクアクセスを継続することが可能となる。

【0053】次に、図13乃至図15を参照して、このディスクアレイ装置15の動作手順を説明する。

【0054】ディスク装置152の故障発生時、アレイコントローラ1511は、故障していないディスク装置152が2つ以上存在するかどうか判定し(図13のステップD1)、存在していれば(図13のステップD1のYES)、有効ブロックテーブルに登録されたデータ

を移行しつつアドレス変換テーブル1514を構成する(図13のステップD2)。

【0055】次に、アレイコントローラ1511は、この移動により有効データの無いディスク装置152が確保できたかどうか判定し(図13のステップD3)、確保できていれば(図13のステップD3のYES)、その有効データの無いディスク装置152上にその他のディスク装置152上のデータに関する冗長データを構成する(図13のステップD4)。

【0056】また、縮退運転中における読み出しリクエスト発生時、アレイコントローラ1511は、そのデータが移動済みかどうかを判定し(図14のステップE1)、移動済みでなければ(図14のステップE1のNO)、要求されたアドレスからデータを読み出す(図14のステップE2)。

【0057】一方、移動済みであれば(図14のステップE1のYES)、アレイコントローラ1511は、アドレス変換テーブル1514を参照して圧縮されたデータの位置を確定し(図13のステップE3)、移動先のアドレスから要求されたデータを読み出してホスト側に引き渡す(図14のステップE4)。

【0058】また、縮退運転中における書き込みリクエスト発生時、アレイコントローラ1511は、そのデータが移動済みかどうかを判定し(図15のステップF1)、移動済みでなければ(図15のステップF1のNO)、要求されたアドレスにデータを書き込む(図15のステップF2)。

【0059】一方、移動済みであれば(図15のステップF1のYES)、アレイコントローラ1511は、この要求されたデータを書き込むだけの空き領域が存在するかどうか判定する(図15のステップF3)。

【0060】ここで、空き領域がなければ(図15のステップF3のNO)、この書き込みリクエストをエラー終了する。一方、空き領域があった場合(図15のステップF3のYES)、アレイコントローラ1511は、アドレス変換テーブル1514を参照してアレイ上の対応する旧データを破棄するとともに、要求されたデータをアレイ上の空き領域に書き込んでアドレス変換テーブル1514を更新する(図15のステップF4)。

【0061】以上のように、このディスクアレイ装置15は、縮退運転中におけるさらなるディスク障害に対しても、データ損失を防ぐことを可能とする。

【0062】なお、前述した実施形態では、故障したディスク装置以外のディスク装置上の有効データを圧縮し、あるいは移動させることにより空きディスク装置を確保する例をそれぞれ説明したが、たとえば有効データを移動させた後にさらに圧縮するようにすれば、空きディスク装置が確保される可能性をより高められることになる。

【0063】

【発明の効果】以上詳述したように、この発明のディスクアレイ装置は、アレイを構成する複数のディスク装置の中のいずれかが故障したことにより縮退運転に移行したときに、(1)その他のディスク装置上の有効データを圧縮し、(2)その他のディスク装置の空き領域が連続するように有効データを移動させ、または、(3)その他のディスク装置の空き領域が連続するように有効データを移動させた後、この移動後の有効データを圧縮する、等を行うことにより、残りのディスク装置の中から空のディスク装置を確保し、この確保した空のディスク装置を用いて、残存する有効データを再生するための新たな冗長データをさらに構築するようにしたことから、縮退運転に移行した際、残存する有効データ(故障したディスク装置上のデータを再生するための冗長データを含む)を再生するための新たな冗長データをさらに構築するため、この縮退運転中にいずれかのディスク装置がさらに故障した場合であっても、新たに構築した冗長データを用いて縮退運転移行時に残存した有効データすべてを保証でき、これにより、通常稼働時のすべてのデータを保証できることになる。

【0064】また、有効データの圧縮前および圧縮後の格納位置の対応関係、または、有効データの再配置前および再配置後の格納位置の対応関係をアドレス変換テーブルに記録し、定期的または所定のイベント発生時に、このアドレス変換テーブルに基づき、その他のディスク装置の空き領域が連続するように、さらには、仮想アドレス順となるように、有効データを再配置することにより、故障したディスク装置上のデータを再生するためのオーバーヘッドを可能な限り小さくする。

【図面の簡単な説明】

【図1】この発明の実施形態に係るディスクアレイ装置を適用するコンピュータシステムの構成図。

【図2】同第1実施形態に係るホストバスアダプタ(HBA)の構成を示す図。

【図3】同第1実施形態のホストバスアダプタ(HBA)のアレイコントローラが行うディスク制御の概要を説明するための図。

【図4】同第1実施形態のホストバスアダプタ(HBA)のアレイコントローラが行うデータ圧縮を説明するための第1の図。

【図5】同第1実施形態のホストバスアダプタ(HBA)のアレイコントローラが行うデータ圧縮を説明するための第2の図。

【図6】同第1実施形態のホストバスアダプタ(HBA)のアレイコントローラが行う圧縮後のデータの再配置を説明するための第1の図。

【図7】同第1実施形態のホストバスアダプタ(HBA)のアレイコントローラが行う圧縮後のデータの再配置を説明するための第2の図。

【図8】同第1実施形態のホストバスアダプタ(HBA)のアレイコントローラの縮退運転時の動作を説明するための第1のフローチャート。

【図9】同第1実施形態のホストバスアダプタ(HBA)のアレイコントローラの縮退運転時の動作を説明するための第2のフローチャート。

【図10】同第1実施形態のホストバスアダプタ(HBA)のアレイコントローラの縮退運転時の動作を説明するための第3のフローチャート。

【図11】同第2実施形態に係るホストバスアダプタ(HBA)の構成を示す図。

【図12】同第2実施形態のホストバスアダプタ(HBA)のアレイコントローラが行うディスク制御の概要を説明するための図。

【図13】同第2実施形態のホストバスアダプタ(HBA)のアレイコントローラの縮退運転時の動作を説明するための第1のフローチャート。

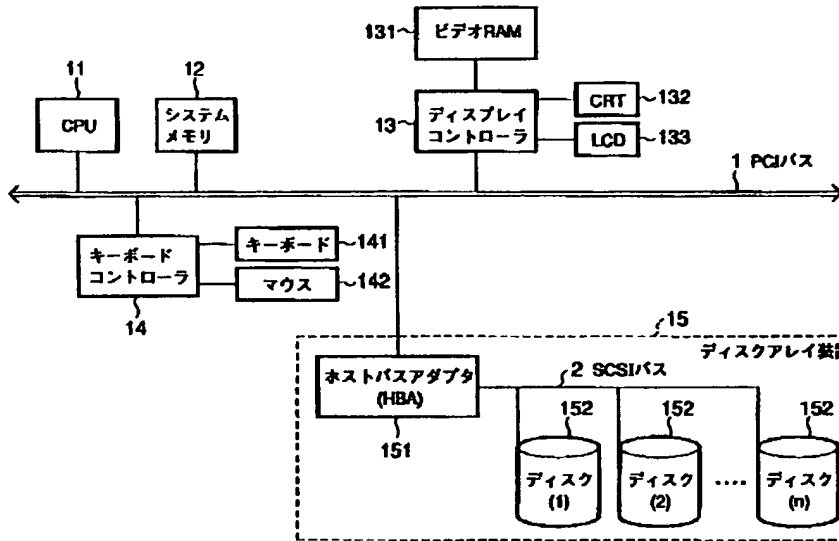
【図14】同第2実施形態のホストバスアダプタ(HBA)のアレイコントローラの縮退運転時の動作を説明するための第2のフローチャート。

【図15】同第2実施形態のホストバスアダプタ(HBA)のアレイコントローラの縮退運転時の動作を説明するための第3のフローチャート。

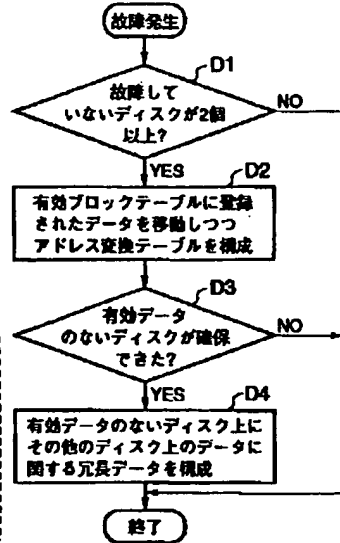
【符号の説明】

- 1...PCIバス
- 2...SCSIバス
- 11...CPU
- 12...システムメモリ
- 13...ディスプレイコントローラ
- 14...動作継続モジュール
- 15...ディスクアレイ装置
- 131...ビデオRAM
- 132...CRT
- 133...LCD
- 141...キーボード
- 142...マウス
- 151...ホストバスアダプタ(HBA)
- 152...ディスク装置
- 1511...アレイコントローラ
- 1512...SCSI制御機構
- 1513...故障回復機構
- 1514...アドレス変換テーブル
- 1515...有効ブロックテーブル

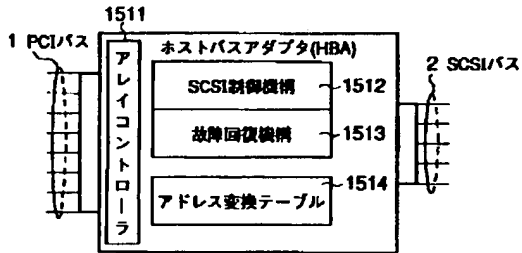
【図1】



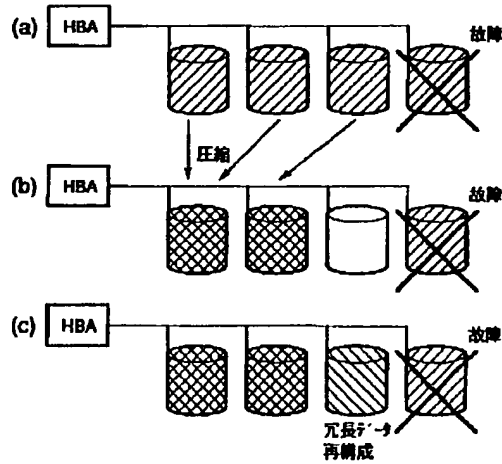
【図13】



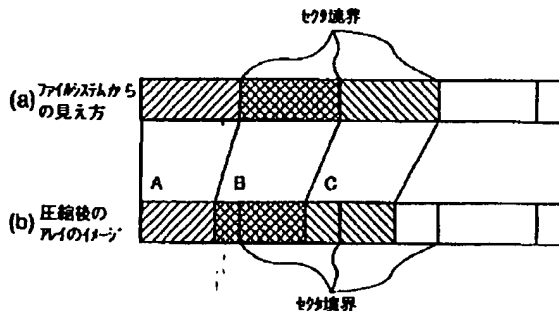
【図2】



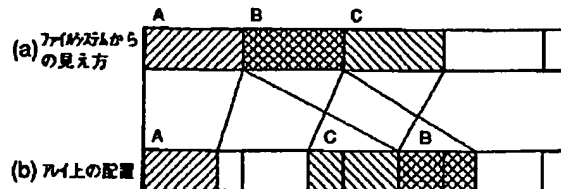
【図3】



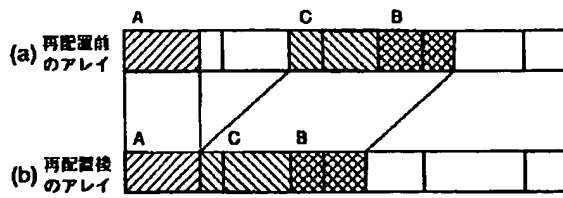
【図4】



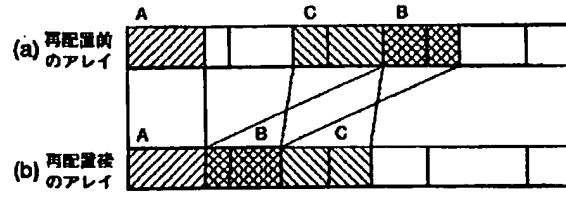
【図5】



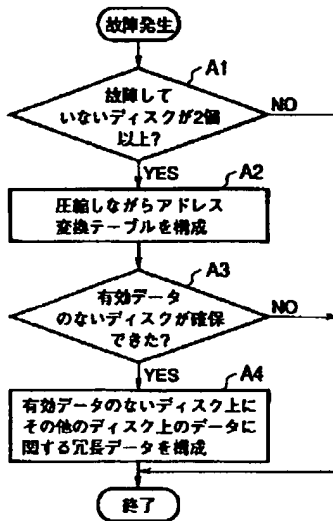
【図6】



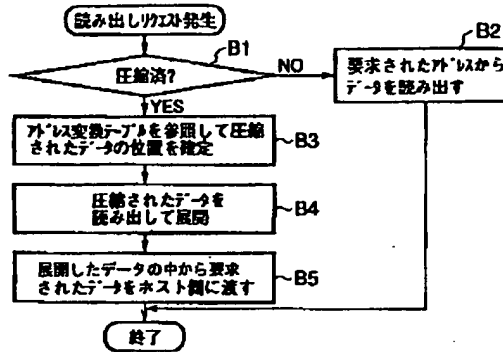
【図7】



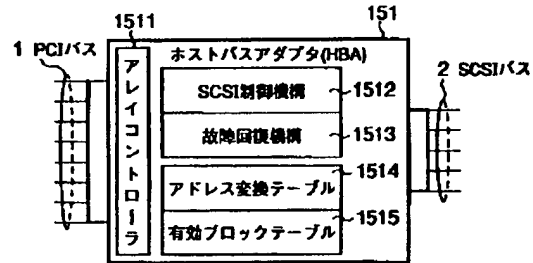
【図8】



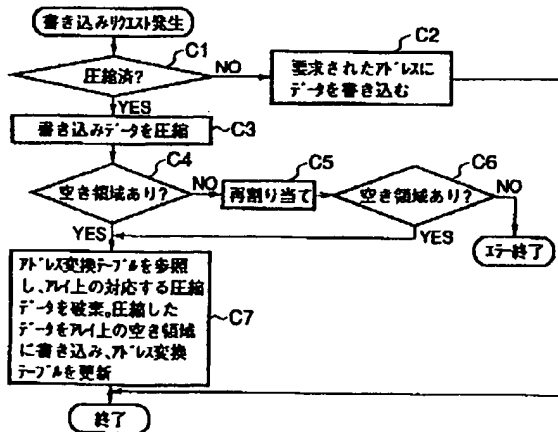
【図9】



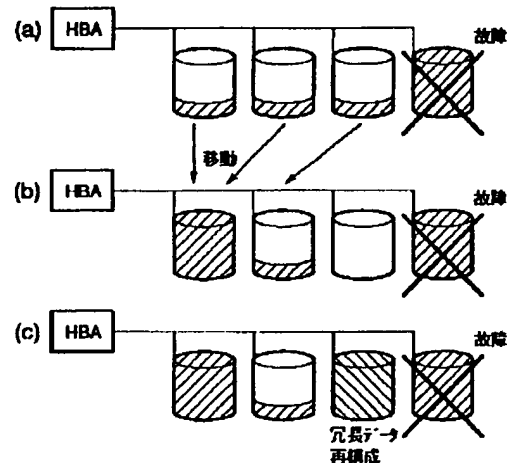
【図11】



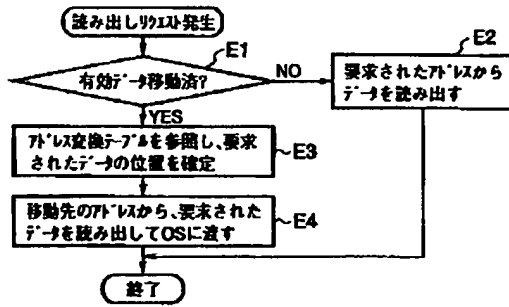
【図10】



【図12】



【図14】



【図15】

